Global Journal of Programming Languages

ENRICHED
PUBLICATIONS

Global Journal of Programming Languages

Aims and Scope

Global Journal of Programming Languages is a peer-reviewed Print + Online journal of Enriched Publications to disseminate the ideas and research findings related to all sub-areas of programming languages. It also intends to promote interdisciplinary researches and studies in computer science maintaining the standard of scientific excellence. This journal provides the platform to the scholars, researchers, and PHD Guides and Students from India and abroad to adduce and discuss current issues in the field of programming and Computer Sciences.

Global Journal of Programming Languages

Global Journal of Programming Languages

(Volume No. 9,   Issue No. 1,   January - April 2024)

Contents

"HADOOP"- A FRAMEWORK IN JAVA

[1]Ms. Ekta Gupta
[1]Java Expert, Brain Mentors Pvt. Ltd.
Email: guptaekta11@gmail.com

A B S T R A C T

Hadoop is an open source project of the Apache Foundation. It is a framework written in Java originally developed by Doug Cutting in 2005. Hadoop uses Google's MapReduce and Google File System technologies as its foundation. It is optimized to handle massive quantities of data which could be structured, unstructured or semi-structured, using commodity hardware, that is, relatively inexpensive computers. This massive parallel processing is done with great performance. Hadoop replicates its data across multiple computers, so that if one goes down, the data is processed on one of the replicated computers. It is a batch operation handling massive quantities of data. As of Hadoop version 2.2, updates are not possible, but appends are possible. Hadoop is used for Big Data. It complements Online Transaction Processing and Online Analytical Processing. It is NOT a replacement for a relational database system.

INTRODUCTION

Early in 2011, Watson, a super computer developed by IBM competed in the popular Question and Answer show Jeopardy. In that contest, Watson was successful in beating the two most winning Jeopardy players. Approximately 200 million pages of text were input using Hadoop to distribute the workload for loading this information into memory. Once the information was loaded, Watson used other technologies for advanced search and analysis. In the telecommunications industry we have China Mobile, a company that built a Hadoop cluster to perform data mining on Call Data Records. China Mobile was producing 5-8TB of these records daily. By using a Hadoop-based system they were able to process 10times as much data as when using their old system, and at one fifth of the cost. In the media we have the NewYork Times which wanted to host on their website all public domain articles from 1851 to 1922. They converted articles from 11 million image files to 1.5TB of PDF documents. This was implemented by one employee who ran a job in 24 hours on a 100-instance Amazon EC2 Hadoop cluster at a very low cost. In the technology field we again have IBM with IBM ES2, an enterprise search technology based on Hadoop, Lucene and Jaql. ES2 is designed to address unique challenges of enterprise search such as the use of an enterprise specific vocabulary, abbreviations and acronyms. ES2 can perform mining tasks to build acronym libraries, regular expression patterns, and geo classification rules. There are also many internet or social network companies using Hadoop such as Yahoo,

Facebook, Amazon, eBay, Twitter, StumbleUpon, Rackspace, Ning, AOL, and so on. Yahoo is, of course, the largest production user with an application running a Hadoop cluster consisting of approximately 10,000 Linux machines. Yahoo is also the largest contributor to the Hadoop open source project. Now, Hadoop is not a magic bullet that solves all kinds of problems. Hadoop is not good to process transactions due to its lack random access. It is not good when the work cannot be parallelized or when there are dependencies within the data, that is, record one must be processed before record two. It is not good for low latency data access. Not good for processing lots of small files although there is work being done in this area, for example, IBM's Adaptive MapReduce. And it is not good for intensive calculations with little data. Big Data solutions are more than just Hadoop. They can integrate analytic solutions to the mix to derive valuable information that can combine structured legacy data with new unstructured data. Big data solutions may also be used to derive information from data in motion. For example, IBM has a product called InfoSphere Streams that can be used to quickly determine customer sentiment for a new product based on Facebook or Twitter comments. Cloud computing has gained a tremendous track in the past few years, and it is a perfect fit for Big Data solutions. Using the cloud, a Hadoop cluster can be setup in minutes, on demand, and it can run for as long as is needed without having to pay for more than what is used.

OPEN SOURCE PROJECTS RELATED TO HADOOP

- Eclipse is a popular IDE donated by IBM to the open source community.
- Lucene is a text search engine library written in Java.
- Hbase is the Hadoop database.
- Hive provides data warehousing tools to extract, transform and load data and then query this data stored in Hadoop files.
- Pig is a high level language that generates MapReduce code to analyze large data sets.
- Jaql is a query language for JavaScript open notation.
- ZooKeeper is a centralized configuration service and naming registry for large distributed systems.
- Avro is a data serialization system.
- UIMA is the architecture for the development, discovery, composition and deployment for the analysis of unstructured data.

MAJOR COMPONENTS OF HADOOP

Hadoop has two major components:

1. **The Distributed File System Component,** the main example of which is the Hadoop Distributed File System, though other file systems, such as IBM GPFS-FPO are supported.

2. **The MapReduce Component,** which is a framework for performing calculations on the data in the distributed file system. Pre-Hadoop 2.2 MapReduce is referred to as MapReduce V1 and has its own built-in resource manager and schedule.

3. **Hadoop Common:** The common utilities that support the other Hadoop modules.

4. **Hadoop YARN:** A framework for job scheduling and cluster resource management.

1. **Hadoop Distributed File System(HDFS) :** HDFS was based on a paper Google published about their Google File System. HDFS runs on top of the existing file systems on each node in a Hadoop cluster. It is not POSIX compliant. It is designed to tolerate high component failure rate through replication of the data. Hadoop works best with very large files. The larger the file, the less time Hadoop spends seeking for the next data location on disk, the more time Hadoop runs at the limit of the bandwidth ofyour disks. Seeks are generally expensive operations that are useful when you onlyneed to analyze a small subset of your dataset. Since Hadoop is designed to run over your entire dataset, it is best to minimize seeks by using large files. Hadoop is designed for streaming or sequential data access rather than random access. Sequential data access means fewer seeks, since Hadoop only seeks to the beginning of each block and begins reading sequentially from there.

2. Hadoop's MapReduce is inspired by a paper Google published on the MapReduce technology. It is designed to process huge datasets for certain kinds of distributable problems using a large number of nodes. A MapReduce program consists of two types of transformations that can be applied to data any number of times - a map transformation and a reduce transformation. A MapReduce job is an executing MapReduce program that is divided into map tasks that run in parallel with each other and reduce tasks that run in parallel with each other.

PIG , HIVE & Jaql – OPEN SOURCE PROJECTS OF HADOOP

Pig, Hive, and Jaql have much in common. They all translate high-level languages into MapReduce jobs so that the programmer can work at a higher level than he or she would when writing MapReduce jobs

in Java or other lower-level languages supported by Hadoop using Hadoop streaming. The high level languages offered by Pig, Hive and Jaql let you write programs that are much smaller than the equivalent Java code. When you find that you need to work at a lower level to accomplish something these high-level languages do not support themselves, you have the option to extend these languages, often by writing user-defined functions in Java. Interoperability can work both ways since programs written in these high-level languages can be imbedded inside other languages as well. Finally, since all these technologies run on top of Hadoop, when they do so, they have the same limitations with respect to random reads and writes and low-latency queries as Hadoop does.

Pig was developed at Yahoo Research around 2006 and moved into the Apache Software Foundation in 2007. Pig's language, called PigLatin, is a data flow language - this is the kind of language in which you program by connecting things together. Pig can operate on complex data structures, even those that can have levels of nesting. Unlike SQL, Pig does not require that the data have a schema, so it is well suited to processing unstructured data. However, Pig can still leverage the value of a schema if you choose to supply one. Like SQL, PigLatin is relationally complete, which means it is at least as powerful as relational algebra. Turing completeness requires looping constructs, an infinite memory model, and conditional constructs. PigLatin is not Turing complete on its own, but is Turing complete when extended with User-Defined Functions.

Jaql was developed at IBM. It is a data flow language like PigLatin but its native data structure format is JavaScript Object Notation, or JSON. Schemas are optional and the Jaql language itself is Turing complete on its own without the need for extension through UDFs. Jaql is a JSON-based query language that, like PigLatin and HiveQL, translates into Hadoop MapReduce jobs. JSON is the data interchange standard that is human-readable like XML but is designed to be lighter-weight. You run Jaql programs using the Jaql shell. You start the Jaql shell using the Jaql shell command.

COMMERCIAL APPLICATIONS OF HADOOP

The HDFS file system is not restricted to MapReduce jobs. It can be used for other applications, many of which are under development at Apache. The list includes the HBasedatabase, the Apache Mahout machine learning system, and the Apache Hive Data Warehouse system. Hadoop can in theory be used for any sort of work that is batch-oriented rather than real-time, is very data-intensive, and benefits from parallel processing of data. It can also be used to complement a real-time system, such as Lambda architecture.

As of October 2009, commercial applications of Hadoop included:

- Log and/or click stream analysis of various kinds
- Marketing analytics
- Machine learning and/or sophisticated data mining
- Image processing
- Processing of XML messages
- Web crawling and/or text processing
- General archiving, including of relational/tabular data, e.g. for compliance

How is Hadoop Different from Past Techniques?

1. **Hadoop can handle data in a very fluid way:** Hadoop is more than just a faster, cheaper database and analytics tool. Unlike databases, Hadoop doesn't insist that you structure your data. Data may be unstructured and schema less. Users can dump their data into the framework without needing to reformat it. By contrast, relational databases require that data be structured and schemas be defined before storing the data.

2. **Hadoop has a simplified programming model:** Hadoop's simplified programming model allows users to quickly write and test distributed systems. Performing computation on large volumes of data has been done before, usually in a distributed setting but writing distributed systems is notoriously hard. By trading away some programming flexibility, Hadoop makes it much easier to write distributed programs.

3. **Hadoop is easy to administer:** Alternative high performance computing (HPC) systems allow programs to run on large collections of computers, but they typically require rigid program configuration and generally require that data be stored on a separate storage area network (SAN) system. Schedulers on HPC clusters require careful administration and since program execution is sensitive to node failure, administration of a Hadoop cluster is much easier.

4. **Hadoop** invisibly handles job control issues such as node failure. If a node fails, Hadoop makes sure the computations are run on other nodes and that data stored on that node are recovered from other nodes.

5. **Hadoop is agile**. Relational databases are good at storing and processing data sets with predefined and rigid data models. For unstructured data, relational databases lack the agility and scalability that is needed. Apache Hadoop makes it possible to cheaply process and analyze huge amounts of both structured and unstructured data together, and to process data without defining all structure ahead of time.

WHY USE APACHE HADOOP?

1. It's cost effective. Apache Hadoop controls costs by storing data more affordably per terabyte than other platforms. Instead of thousands to tens of thousands per terabyte, Hadoop delivers compute and storage for hundreds of dollars per terabyte.

2. It's fault-tolerant. Fault tolerance is one of the most important advantages of using Hadoop. Even if individual nodes experience high rates of failure when running jobs on a large cluster, data is replicated across a cluster so that it can be recovered easily in the face of disk, node or rack failures.

3. It's flexible. The flexible way that data is stored in Apache Hadoop is one of its biggest assets – enabling businesses to generate value from data that was previously considered too expensive to be stored and processed in traditional databases. With Hadoop, you can use all types of data, both structured and unstructured, to extract more meaningful business insights from more of your data.

4. It's scalable. Hadoop is a highly scalable storage platform, because it can store and distribute very large data sets across clusters of hundreds of inexpensive servers operating in parallel. The problem with traditional relational database management systems (RDBMS) is that they can't scale to process massive volumes of data.

REFERENCES

1. http://bigdatauniversity.com
2. http://www.cloudera.com/content/cloudera/en/about/hadoop-and-big-data.html
3. http://hadoop.apache.org
4. http://hortonworks.com/hadoop/

5. *http://www-01.ibm.com/software/data/infosphere/hadoop/*

6. *https://www.mapr.com/products/apache-hadoop*

7. *http://www.sas.com/en_us/insights/big-data/hadoop.html*

8. *https://www.oracle.com*

9. *http://www.sas.com/enus/insights/big-data/hadoop.html*

10. *"Hadoop-related projects at"* . *Hadoop.apache.org. Retrieved 2013-10-17.*

11. *"Hadoop"* *Azure.microsoft.com. Retrieved 2014-07-22.*

12. *"HDFS Architecture". Retrieved 1 September 2013.*

13 *Murthy, Arun (2012-08-15). −"Apache Hadoop YARN Concepts and Applications".hortonworks.com. Hortonworks. Retrieved 2014-09-30 "What is the Hadoop Distributed File System (HDFS)?" . ibm.com. IBM. Retrieved 2014-10-30.*

Study of Software Test box Generation Techniques

[1]Mohit Saifi

[1,] M. Tech scholar, Computer Science and Engineering

Al-Falah School of Engineering and Technology. Dhouj, Faridabad, Haryana

A B S T R A C T

Software testing is an essential element of software development which sure the verification and validation method of the software. In arrange to do software testing we necessity have to apply the method of mapping the software for all its transition states and alone validating the output for a set of known input. For a any given part of software we will be writing a set of test cases that called test suites and it is used to collection together similar test cases. Test suites is a collection of test cases that are planned to be used to test a software program to demonstrate that it has some specific set of behaviors. In order to find out how a test case is valid or not for that we do not have specific device. We mostly depend on the software testers understanding of the obligation. The scope of this paper to study different technique use in test box, for example test box generation using genetic algorithm, test case generation using random based testing, test case generation using Model based testing. The test cases are derived by analyzing the dynamic behavior of the objects due to external and internal stimuli.

Keywords:Software Testing, Model based testing, Random Testing, Test box, Genetic algorithm

I. INTRODUCTION

Software testing is a significant movement in software development life cycle. Software organizations spend large segment of their budget in testing related behavior. Software Testing is a costly and time overriding process in software development life cycle. Automation of this phase may lead to overcome the above problems and also reduces the human effort in other ways it also helps in detecting the human intended errors and logical errors as well. The Automation of testing will not be that much productive in terms of cost and time consuming because if we have to wait till the end of the Software Development Life Cycle stage (SDLC). Software Testing[1] is any process or activity aimed at evaluating a system or attribute or capability of a program and determining through the purpose to find that whether it satisfies or meets the specified requirements or not. In simple words testing is executing a system in order to discover any errors, gaps or missing requirements in contrary to the actual requirements desire output.
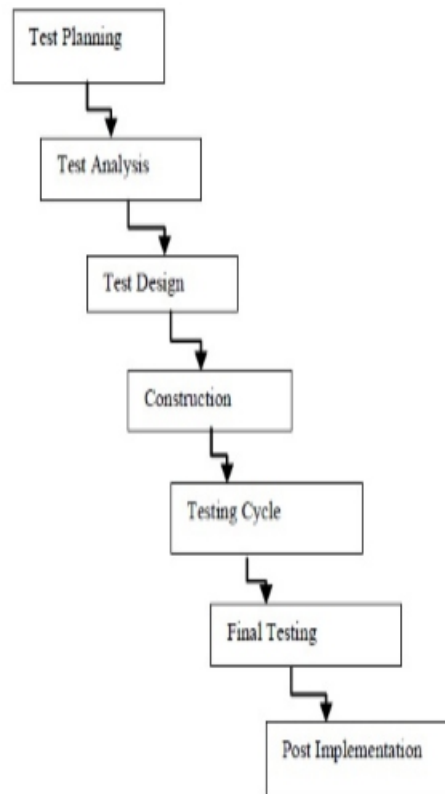
```
Test Planning
   │
   ▼
Test Analysis
   │
   ▼
Test Design
   │
   ▼
Construction
   │
   ▼
Testing Cycle
   │
   ▼
Final Testing
   │
   ▼
Post Implementation
```

Figure1: Software Testing Steps

Test cases help the user to pen down entire coverage to the application and test all possible combinations in the application. It also provides the user to easily reproduce the steps that were undertaken to uncover a defect that as detected during test. It also provides the extent of which the testing has concluded and the areas in which the application is working fine. Throughout the years a huge number of different technique have been proposed for generating test cases. A test case is basically a description of a test. A Test case has the components that describe an input, event or action and expected response to determine if the feature of an application is working correctly. Test cases can be mapped directly to. Test cases are derived from use cases and can also be derived from system requirements. The one of the main advantages of generating test cases from requirements specifications and design will often help the software or test engineer to discover problems as early. As they can be created earlier in the development life cycle and get ready for use before the programs are constructed. Generating test cases early helps Software Engineers or test engineer can often find ambiguities and inconsistencies in the requirements specification and design documents. This will absolutely take down the cost of building the software systems as errors are eliminated early during the life cycle. The below diagram[2] highlights the steps of software testing.

II. TESTING STAGES

Test cases assist the user to pen down entire reporting to the application and test all possible combinations in the application. It also provides the user to easily reproduce the steps that were undertaken to uncover a defect that as detected during test. It also provides the extent of which the testing has concluded and the areas in which the application is working fine. Through the years a number of different methods have been proposed for generating test cases. A test case is a explanation of a test, independent of the way a given system is designed. Test cases can be mapped directly and derived from use cases. Test cases can also be derived from system requirements.. Additionally, when the test cases are generated early, Software Engineers can often find ambiguities and inconsistencies in the requirements specification and design documents. This will definitely get down the cost of building the software systems as errors are eliminated early during the life cycle. The below diagram[13] highlights the steps of software testing.

The above steps are under taken to test an application or software. Firstly the test plan is created which includes the overall structure of the plan i.e. requirements for testing, testing flow etc. Then the test cases are written according to some specific technique. There are many existing techniques are available. These cases are executed under suitable environment. The bugs are found out and reported to the developers. The developer fixes these bugs and finally these are verified using some testing technique. This way the whole procedure works.
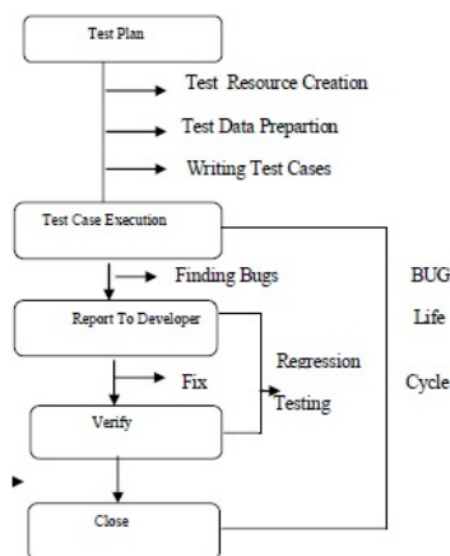
Figure 2 : Software Testing Stages

III. SOFTWARE TESTING METHOD

Method1: Test Case generation using Genetic Algorithms (GA): In this technique, automated generation of test cases in object oriented systems has been presented. The test cases are derived by analyzing the dynamic behavior of the objects due to internal and external stimuli. The study is limited to the object diagrams taken from the UML (Unified Modeling Language) model of the system. In order to carry out all suitable test cases of a given object diagram, Genetic Algorithm's (GA) tree crossover has been proposed.

Method2: Test Case generation using Random Testing : Random test case generation is a technique where in the test cases are generated not based on an algorithm but based on the ones assumption of the application. This technique is implemented on the above mentioned case study. The following classes will be tested and various test inputs will be provided to check for the faults. The framework used here to validate is called AutoTest and using this we will be able to predict the number of issues found and the number of detects undetected. The AutoTest framework classifies test cases into the following categories: passed (no exception), unresolved (precondition violation in method under test), or failed (other exception). Method create test depicted below contains the main loop of the testing strategy used in this technique. At each step it selects a method for testing and then causes the execution of this method.

```
create_test ( timeout ) :
from initialize pool
until timeout
loop
 m := choose ( methods under test () )
create_test for method (m)
end
```

1. Method Initialize pool creates an empty pool of objects to be used for testing .
2. The Method methods under test returns the set of methods under test.
3. The non-deterministic method choose selects an arbitrary element of a set or a list.
4. The method create test for method is responsible for generating a call to method m.

Method3: Test box generation using a combination of Activity Diagrams and Sequence diagrams
Category Partition Method is simply a specification based testing technique with respect to some specific criteria. CPM first decomposes the functional specification into functional units and then examines each functional unit. It finds the categories for each parameter and environmental condition. It

helps in identifying the parameters and environmental conditions that affect the execution behavior of the function. CPM is also helpful in finding the categories of information that characterize each parameter and environmental condition. We will first generate test scenarios from activity diagrams, which achieve path coverage criteria perfectly, followed by generation of test cases by analyzing the respective sequence and class diagrams of each scenario. This technique helps to reuse the design. Complex tests are built up by designing a test that runs through a series of use cases. First the test scenarios will be generated followed by analysis of each scenario using the sequence diagrams to formulate the test cases. Following gives a pictorial representation of the approach followed.

Method4: Test Case generation using Model Based Testing Model Based Testing (MBT) is a black-box testing technique where common testing tasks such as test case generation and test result evaluation are automated based on a model of the application under test. This approach has recently spread to a variety of software domains but originates from hardware testing, most notably from telephone switches, and from the increasing use of object orientation and models in software design and software development.
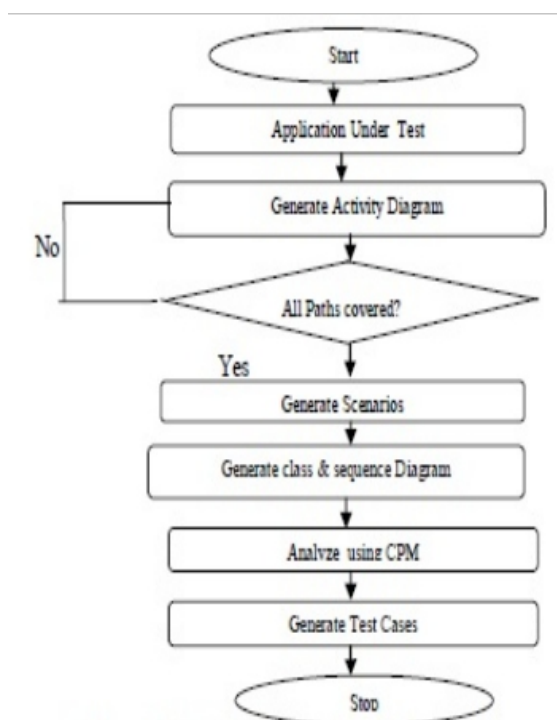
Figure 3. Pictorial Representation of Test scenario

Model Based Testing (MBT) automate the complete design of test cases and the generation of the traceability matrix, which traces the link between requirements and generated test cases. Instead of writing hundreds of test cases, the test designer constructs an abstract model of the system under test. The MBT tool is used to generate a set of test cases from that model. Below describes the manner in which the MBT takes place.

Figure4. Model based Testing Process Diagram

IV.   CONCLUSION

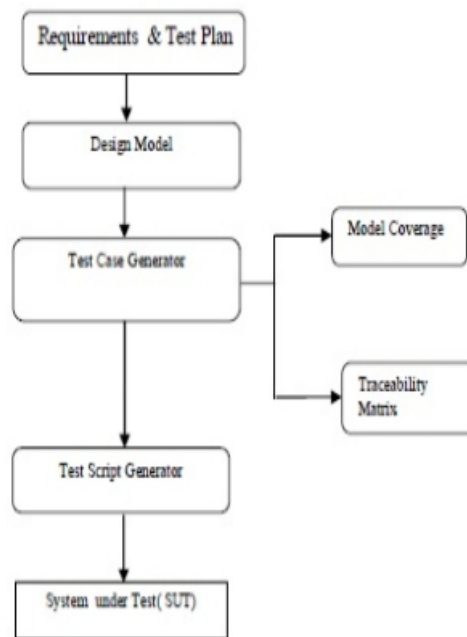We create various test box technique in a straight line from UML behavioral diagram, where the design is reused. By using our move toward defects in the design model can be detected during the study of the model itself. So, the defects can be removed as early as possible, thus reducing the cost of defect removal. First we generate test case generation using genetic algorithm, random based testing, model based testing and test scenarios from the activity diagram and then for each scenario the corresponding sequence diagram generated .After analyzing each category, its significant values and constraints are generated and respective test cases are derived. Test coverage criteria achieved is another advantage of our approach.

REFERENCES

1. Aljahdali S.H , Ghiduk A.S and Telbemy M.(2010), The limitations of Genetic algorithms in software testing, International Conference on Computer Systems and Applications, held at Hammamet from 16th May to 19th May 2010, ISBN: 978-1-4244-7716-6, pp: 1-7..

[2]. Jin-Cherng Lin and Pu-Lin Yeh, (2001), Automatic Test Data Generation for Path Testing Using Genetic Algorithms, International Journal of Information Sciences, January 2001, New York, USA, Volume 131, Issue 1-4,

[3]. Rudnick Elizabeth M., Greenstein Gary S., "A Genetic Algorithm Framework for Test Generation" IEEE Transactions on Computer Aided design of Integrated Circuits and Systems, VOL. 16, NO. 9, 1997.

[4]. R.P. Pargas, M.J. Harrold, and R.R. Peck, "Test data generation using genetic algorithms", The Journal of Software Testing, Verification and Reliability, Vol.9, No.4, Pp. 263 – 282, 1999

[5]. T.Y. Chen, Pak-Lok Poon and T.H Tse," A choice Relation Framework for supporting category-partition Test case Generation", IEEE Transactions on software engineering Vol. 29, No.7, 2003.

[6]Ghiduk, Ahmed S and Girgis, Moheb R.(2010), Using Genetic Algorithms and Dominance Concepts for Generating Reduced Test Data, Informatica (Slovenia), Volume 34, Number 3, pp.377-385.

[7]. Sarma Monalisa and Mall Rajib, "Automatic Test Case Generation from UML Models," 10th International Conference on Information Technology, pp. 196-201, 2007.

[8]. Ilinca Ciupa, Alexander Pretschner, Andreas Leitner, Manuel Oriol, Bertrand Meye: "On the Predictability of Random Tests for Object-Oriented Software" International Conference on Software Testing, Verification, and Validation, 2008.

[9]. Reza, H., K. Ogaard and A. Malge," A model based testing technique to test web applications using statecharts", Proceedings of 5th International Conference on Information Technology: New Generations, pp. 183-188, 2008.

[10]. Prasanna M., Chandran K.R., "Automatic Test Case Generation for UML Object diagrams using Genetic Algorithm",Vol.1,No.1,2009.

[11]. Singh Kulvinder and Kumar Rakesh, "Optimization of Functional Testing using Genetic Algorithms" International Journal of Innovation, Management and Technology, Vol. 1,No. 1, 2010.

[12] Sthamer "The automatic generation of software test data using genetic algorithms". Phd thesis, university of Glamorgan, Pontyprid, wales, Great Britain, 1996.

Soft Computing Methods For Web Intelligence

[1]Nishat Zulquernain, [2]MD Waseem Akram

[1,] M.Tech (Scholar)

Department of computer science

Al-Falah School of Engineering & Technology

Dhauj Faridabad

nishat.cse.09@gmail.com

[2,] MD Waseem Akram

M.Tech (Scholar)

Department of computer science

Al-Falah School of Engineering & Technology

Dhauj Faridabad

wwww.waseem@gmail.com

A B S T R A C T

Web has become the primary means for information distribution. It is being used for educational purposes and thus its popularity resulted in heavy traffic in the Internet. Web Intelligence (WI) deals with the scientific exploration of the new territories of the Web. As a new field of computer science, it combines artificial intelligence and advanced information technology in the context of the Web, and goes beyond each of them. Data mining has a lot of scope in e-Applications. The key problem is how to find useful hidden patterns for better application. Problem to address soft computing techniques like Neural networks, Fuzzy Logic, Support Vector Machines, Genetic Algorithms in Evolutionary Computation. In this paper, we explore soft computing techniques use to achieve web intelligences.

Keywords: Soft Computing, Neural Network, Data Mining, Web Intelligence

I. INTRODUCTION

Data mining has useful business applications such as finding useful hidden information from databases, predicting future trends, and making good business decisions [1,6,7]. Soft computing techniques such as fuzzy logic, genetic algorithm and neural networks are useful in data mining [l,2,l0]. Web intelligence, a term that was coined in the late 1999's, concerns about research and application of machine learning and information technology with a specific focus on the Web platforms. Typical Web Intelligence applications include but not limited to online text classification, Web document clustering, Web recommender for e-commerce, Web usage profiling and similar knowledge discovery tasks are drawing attention from communities of global researchers. The data, in the context of data that are originated from the Web, called Web Intelligence data pose certain challenges to knowledge discovery

tasks and Web mining. WI (Web Intelligence) is studied carefully from different aspects [3]. WI exploits Artificial Intelligence (AI) and advanced Information Technology (IT) on the Web and Internet [3].

1)Web- The World Wide Web (abbreviated as WWW or W3, commonly known as the web) is a system of interlinked hypertext documents accessed via the internet With a we, one can view web pages that may contain text, images, videos, and other multimedia and navigate between them via.hyper link On March 12,1989,Tim berners lee a british computer scientist and former cern(European Organization for Nuclear Research) employee, wrote a proposal for what would eventually become the World Wide Web.[5].

2)Web Intelligence- Web intelligence is the area of study and research of the application of artificial intelligence and information technology on the web in order to create the next generation of products, services and frameworks based on The internet.The term was born in a paper written by Ning Zhong, Jiming Liu Yao and Y.Y.Ohsuga in the Computer Software and Applications Conference in 2000. [11]

3) Methods of Data Mining – .Artificial neural networks - Non-linear predictive models that learn through training and resemble biological neural networks in structure. Warren McCulloch and Walter Pitts [19] (1943) created a computational model for neural networks based on mathematics and algorithms. They called this model threshold logic. Neural network is used in data mining for pattern recognition.

Decision trees - Tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset. Although decision trees have been in development and use for over 60 years (one of the earliest uses of decision trees was in the study of television broadcasting by Belson in 1956).Decision tree is used in data mining for the classification.

A decision tree consists of 3 types of nodes:
a. Decision nodes - commonly represented by squares.
b. Chance nodes - represented by circles.
c. End nodes - represented by triangles.

Rule induction - The extraction of useful if-then rules from data based on statistical significance. The rule induction algorithm was first used by Hunt in his CLS system in 1962.
Genetic algorithms - Optimization techniques based on the concepts of genetic combination, mutation, and natural selection. It was introduced by John Holland in 1975.In 1989, Axcelis, Inc.

Nearest neighbour - A classification technique that classifies each record based on the records most similar to it in an historical database. Donald Knuth in vol. 3 of The Art of Computer Programming (1973) called it the postoffice problem, referring to an application of assigning to a residence the nearest post office. Nearest is used in data mining for clustering.

II. Literature review

Soft Computing-

Soft computing is an emerging approach to computing which parallels remarkable ability of human mind to reason and learn in an environment of uncertainty and imprecision[12].Soft Computing consists of several computing paradigms like Neural Networks, Fuzzy Logic, and Genetic algorithms. Soft Computing uses hybridization of these techniques. A hybrid technique would inherit all the advantages of constituent techniques. Thus the components of Soft Computing are complementary, not competitive, offering their own advantages and techniques to partnerships to allow solutions to otherwise unsolvable problems [13].

1. Difference between soft and hard computing-

Hard computing	Soft computing
Conventional computing	Soft computing is toleran to fimprecision.
Requires a precisely stated analytical model.	
Often requires a lot of computation time.	Can solve some real world problems in reasonably less time.
Not suited for real world	Suitable for real world
problems for which ideal	problems
model is not present.	
It requires full truth.	Can work with partial
	truth.
It is precise and accurate.	Imprecise.
High cost for solution.	Low cost for solution

2. Soft Computing Methodsa)

a) Fuzzy Logic- As one of the principal constituents of soft computing, fuzzy logic is playing a key role in what might be called high MIQ (machine intelligence quotient) systems. Two concepts within fuzzy logic play a central role in its applications. The first is a linguistic variable; that is, a variable whose values are words or sentences in a natural or synthetic language[14]. The other is a fuzzy if-then rule, in which the antecedent and consequents are propositions containing linguistic variables [14]. While variables in mathematics usually take numerical values, in fuzzy logic applications, the non-numeric

linguistic variables are often used to facilitate the expression of rules and facts [15]. For example, a simple temperature regulator that uses a fan might look like this:

IF temperature IS very cold THEN stop fan
IF temperature IS cold THEN turn down fan
IF temperature IS normal THEN maintain level
IF temperature IS hot THEN speed up fan

There is no "ELSE" – all of the rules are evaluated, because the temperature might be "cold" and "normal" at the same time to different degrees.

b) Neural networks- Based on the computational simplicity Artificial Neural Network (ANN) based classifier are used. In this proposed system, a feed forward multilayer network is used. Back propagation (BPN) Algorithm is used for training. There must be input layer, at least one hidden layer and output layer. The hidden and output layer nodes adjust the weights value depending on the error in classification. In BPN the signal flow will be in feed forward direction, but the error is back propagated and weights are up dated to reduce error. The modification of the weights is according to the gradient of the error curve, which points in the direction to the local minimum. Thus making it much reliable in prediction as well as classifying tasks.
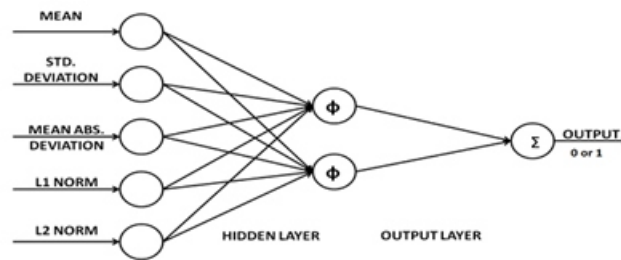

Fig Artificial Neural Network Structure

In propose Methodology, We train an ANN with back propagation as error tracing method and weight updated with biological genetic algorithm. Proposed Method consist of two phase.

c) Support Vector Machines- Support Vector Machine (SVM) is a classification technique based on statistical learning theory. It is based on the idea of a hyper plane classifier. The goal of SVM is to find a linear optimal hyper plane so that the margin of separation between the two classes is maximized [17].The figure 2 below presents an overview of the SVM process
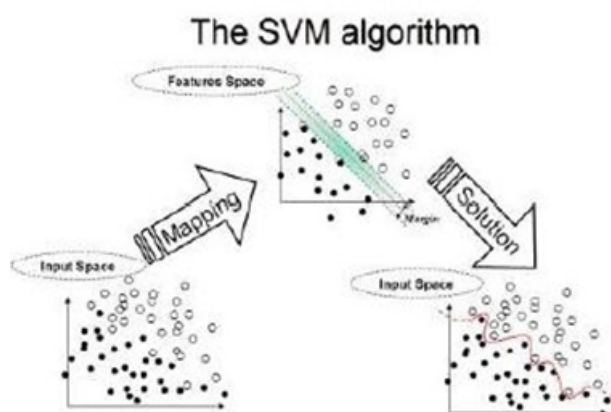
The SVM algorithm

Fig SVM process

Genetic Algorithm- A genetic algorithm (GA) is a search heuristic that mimics the process of natural selection. This heuristic (also sometimes called a metaheuristic) is routinely used to generate useful solutions to optimization and search problems.[16]

1. In proposed Methodology, GA work as

2. Parent selection Previous weight applied to error occurred in NN layer are the parent chromosome.

3. Split parent into two parents.

4. Select Crossover indices from parent then swap the element to generate new offspring.

5. Combine this two offspring and train NN with newly generated child.

6. Set weights to layer at which the minimum error occurred.

7. Muted selected indices

8. Repeat from step no 1 to step no 6 for every error occur NN Layers.

9. Stop.

2.3 Web Intelligence- This definition has the following implications. The basis of WI is AI and IT. The "I" happens to be shared by both "AI" and "IT", although with different meanings in them, and "W" defines the platform on which WI research is carried out, [18]. The goal of WI is the joint goals of AI and IT on the new platform of the Web.

1. Web Intelligence Ralated Topics-

An overview of WI related topics as shown in Figure 3 and list several major subtopics in each topic below [11].
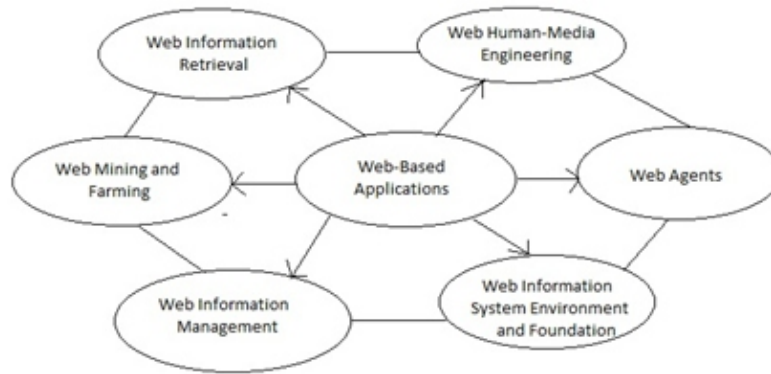
Fig-WI topics

a. Web Human-Media Engineering: The art of Web page design, multimedia information representation, multimedia information processing, visualization of Web information, and Web-based human computer interface.

b. Web Information Management: Data quality management, information transformation, Internet and Web-baseddata management, multidimensional Web databases and OLAP (on-line analytical processing), multimedia information management, new data models for the Web, object oriented Web information management, personalized information management, semi-structured data management, use and management of metadata, Web knowledge management, Web page automatic generation and updating, as well as Web security, integrity, privacy and trust.

c. Web Information Retrieval: Approximate retrieval, conceptual information extraction, image retrieval, multilinguistic information retrieval, multimedia retrieval, new retrieval models, ontology-based information retrieval, as well as automatic Web content cataloging and indexing.

d. Web Agents: Dynamics of information sources, e-mail filtering, email semi-automatic reply, global information collecting, information filtering, navigation guides, recommender systems, remembrance agents, reputation mechanisms, resource intermediary and coordination mechanisms, as well as Web-based cooperative problem solving.

e. Web Mining and Farming: Data mining and knowledge discovery, hypertext analysis and transformation, lerning user profiles, multimedia data mining, regularities in Web surfing and Internet congestion, text mining, Web-based ontology engineering, Web-based reverse engineering, Web farming, Web-log mining, and Web warehousing.

f. Web Information System Environment and Foundations: Competitive dynamics of Web sites, emerging Web technology, network community formation and support, new Web information description and query languages, theories of small world Web, Web information system development tools, and Web protocols.

III. CONCLUSIONS

Current research in soft computing mainly focuses on the discovery algorithm and visualization techniques. Soft computing methodologies, like fuzzy logic, neural networks, Support Vector Machines and genetic algorithms have recently been used to solve data mining problems can also be in web intelligence. This paper suggest the soft computing techniques used in soft computing. according to the new characteristic, so to achieve intelligence of the web.

REFERENCES

[1] K. Cios, W. Pedrycz, and R. Swiniarski, Data Miming methodr for Knowledge Discovery, Kluwer Academic Publishers, 1998.

[2] Meijuan Gao, Jingwen Tian, and Shiru Zhou ,"Research of Web Classification Mining Based on Classify Support Vector Machine" 2009 ISECS International Colloquium on Computing, Communication, Control, and Management

[3] Y.Y. Yao, Zhong, N., Liu, J. and Ohsuga, S., "Web Intelligence (WO:Research challenges and trends in the new information age", Proc. Of WI2001, pp. 1-17, Springer, 2001.

[4] Y.-Q. Zhang and T.Y. Lin, "Computational Web Intelligence (CWI): Synergy of Computational Intelligence and Web Technology," Proc. of FUZZIEEE2002 of World Congress on Computational Intelligence 2002: Special Session on Computational Web Intelligence, pp.1104-1107, May 2002.

[5] "World Wide Web Consortium". "The World Wide Web Consortium (W3C)..."

[6] K. Thearling, "Increasing customer value by integrating Data Mining and Campaign Management SoAware," Direct Marketing Magazine, February 1999.

[7] C. Westpbal and T. Blaxton, Data Mining Solulions -Methods and Tools for Solving Real- World Problems,John Wiley \& Sons, 1998.E. Gelenbe, Y. Feng, K.R.R. Krishnan, Neural Network Methods for Volumetric Magnetic

[8] V.Vapnik, Statistical Learning Theory, Wiley, 1998.

[9] J.Weston and C. Watkins, Support vector machines for multi-class pattern recognition, In Proceeding of the 6th European symposium on Artificial Neural Network (ESANN), 1999.

[10] Y,-Q. Zhang, M. D. Fraser, R. A. Gagliano and A.Kandel, "Granular Neural Networks for Numerical-Linguistic Data Fusion and Knowledge Discovery,"Special Issue on Neural Networks for Data Mining and Knowledge Discovery, IEEE Transactions on Neural Nehvorks, Vol. 11, No. 3, pp.658-667, May, 2000.

[11] Zhong, Ning; Liu Yao, Jiming; Yao, Y.Y.; Ohsuga, S. (2000), "Web Intelligence (WI)", Web Intelligence, Computer Software and Applications Conference, 2000. COMPSAC 2000. The 24th Annual International, p. 469, doi:10.1109/CMPSAC.2000.884768, ISBN 0-7695-0792-1.

[12] J. S. R. Jang, C. T. Sun and E. Mizutani, Neuro-Fuzzy and Soft Computing :A Computational Approach to Learning & Machine Intelligence, Prentice Hall, Saddle River, New Jersey, 1997.

[13] S.N. Sivanandam, S.N. Deepa, Principles of Soft Computing, Wiley India (P) Ltd, 2007.

[14] L.A. Zadeh, "Outline of a New Approach to the Analysis of Complex Systems and Dccision Processes," IEEE Trans. Systems, Man and Cybmzetics, pp. 28-44.

[15] Zadeh, L. A. et al. 1996 Fuzzy Sets, Fuzzy Logic, Fuzzy Systems, World Scientific Press, ISBN 981-02-2421-4.

[16] Mitchell, Melanie (1996). An Introduction to Genetic Algorithms. Cambridge, MA: MIT Press. ISBN 9780585030944.

[17] C.J.C. Burges, " A tutorial on support vector machines for pattern recognitation " , Data Mining and Knowledge Discovery , vol.2 , 1998, pp. 121-167.

[18]] Y.Y. Yao, "Web Intelligence (WI): New Frontiers Exploration" University of Regina, Saskatchewan, Canada 2003.

[19] McCulloch, Warren; Walter Pitts (1943). "A Logical Calculus of Ideas Immanent in Nervous Activity". Bulletin of Mathematical Biophysics 5 (4): 115–133. doi:10.1007/BF02478259.

[20] Markoff, John (29 August 1990). "What's the Best Answer? It's Survival of the Fittest". New York Times. Retrieved 9 August 2009.

[21] Yuehui Chen, "Hybrid Soft Computing Approach to Identification and Control Of Nonlinear Systems".

[22] Xiaozhe Wang, Ajith Abraham and Kate A. Smith, "Soft Computing Paradigms for Web Access Pattern Analysis"

A New Approach Towards Distributed File System

[1]Sangeeta Lalwani, [2]Subhash Chand Gupta
[1]M.Tech (CS&E),
Amity University, Noida, India.
E-mail: Sangeeta.lalwani29@gmail.com
[2]Assistant Professor (CS&E),

Amity University, Noida, India.

E-mail: , scgupta@gmail.com

A B S T R A C T

A Java Distributed File System is an implementation of classical distributed time sharing model which is constructed on a file system where it allows one or many users to share the files and storage resources safely and securely. The system is an effective, safe and secured way of keeping the data which is private or needed to be kept safe. File replication, naming and transparency etc such techniques are already been used in other existing systems. An existing Distributed File System must be having File Replication technique but it might not encrypt the file. Some other Distributed File System might encrypt the files but have bad file replication. So here is the use of Java Distributed File System where all these different features are combined together in one. The Distributed File System will be written in Java because Java is a powerful platform for advancement of distributed systems.

Keywords- Naming and Transparency, File Replication, Fault Tolerance, Effective

INTRODUCTION

A file system is the method of permanent storage of files and organizing those files so that it is easy to find those files and access them. A distributed file system is conversely a network of file systems which provides the storage of the files among the commodity of connected computers. A typical configuration of Distributed File System is a collection of workstations and mainframes connected by a local area network. A Distributed File System is implemented as part of the operating system of each of the connected computers. One of the main drawbacks found in the prevailing file systems is the lack of the scalability. Security is quickly becoming a mandatory feature of data storage systems (Park, 2004). A secure storage system should protect the confidentiality and the integrity of the stored data. Unauthorized access may occur from other nodes on the network. Data that resides on the storage device, an attacker may physically access the storage device or send appropriate commands over the network. If the network is not secure, these commands may also be initiated by clients that are

authorized to access other parts of the networked storage system. Some of the key features of the system are Manipulation of files using framework, improved scalability and improved reliability, security, authentication and high performance.

The system supports the file sharing through intranet. A file service which is very well designed provides access to the stored files at a server with the performance and reliability similar to that of the disk. In other words it can be said that a Distributed File System allows the users to store and access remote files like in the local way, but from any computer in the intranet. It has a capability of distributing the files from one computer to other computers. The system supports the common operations effectively on the same kind of sharing resources. A Distributed Operating System can be said as a collection of interconnected computers which appears to the users or the clients as a single system. If any of the system does not work or gets crashed then it would not affect the working of the other systems interconnected. It manages the data that is the files and migrate the data or file from one site to other sites. The system supports the replication of files that improves the availability and performance. The system also employs asynchronous communication that is more efficient when there is low loss and low error rates over the transmission medium because data is not retransmitted and no time is spent setting negotiating the connection parameters at the beginning of the transmission. Also the system performs a task by a set of rules that controls the data or files communication. In this case when the client opens the files the distributed application request is translated into set of protocol messages. In order to fulfil the request by the client, a response is generated by the server to each protocol. In other simple words the protocol provides the communication between the client and server. It plays a vital role in effective communication.

LITERATURE REVIEW

An extensive review of research was carried out in order to gain better understanding of Distributed File System and the tools and the technical skills required to implement the user's requirement. There are several layers upon which a distributed application is built. Starting from the lowest there is a network through which a group of host computers are connected together so that they can communicate with each other. TCP/IP is a protocol which let the computers to send the data over the network to the other connected computers. Security protocols and directory services are the higher level services that can be defined on the top of the networked protocol. Using the mid-level services and network protocols, the distributed applications runs at the top of these layers. Process, Threads, Objects, Agents are the further broken down parts of the distributed application. First talking about the process, a process is something which is created describing a sequence of the steps during and compilation of the code and when the code is been run on the computer. A process is responsible for performing the tasks like receiving and

sending of the data or files over the network. The main purpose of a thread is to provide feedback through the output devices for example monitor. There has to be at least one thread of control in a process and a thread can independently run from other threads. The object-orientated programs are made up of objects which can be accessed by one or more threads within the process. The processes which are created in the distributed application could be made up of one or more than one objects. Last but not the least part is the agents which are the high level system component and they usually are the functional elements of a distributed application[5]. Although Distributed File System supports the network wide sharing of files but then also there are various issues that are encountered in present systems such as file sharing, naming and transparency, cache consistency problems, updated file replications. Sun's Network File System that is being used in the businesses today is mainly deployed for Local Area Network. The various advantages offered are scalability and transparency. The mapping among logical and physical objects causes the abstraction of file. Through this the location of file gets hidden that is termed as file transparency. Java Based Distributed File System provides high scalability as it supports file transparency that is not yet provided by present Distributed File System. Although the network file system offers various advantages but it also involves various disadvantages such as migration transparency that is not supported through it. Also it does not support file replication with updations due to which a client have to suffer heavy loss of data in certain situations. Secondly, a most popular distributed system is Hadoop Distributed File System. The system is cost effective as it does not require expensive hardware. The various disadvantages of system are that it does not cover all the requirements. It is originally designed for input-output bounds. Therefore it is more time consuming and non-conservative in CPU and memory.

Java Based Distributed File System comes upon as a new way or as a new challenge in solving the faults and problems that occur in the present file distributed systems.Many different concepts such as transparency, fault tolerance, scalability have been defined and solved using this approach. Alternatives for providing access and semantics of sharing have also been discussed. The system provides various functionalities and plays a vital role in overcoming the difficulties of distributed file systems being used today. Some of them are effective communication through asynchronous communication, cache consistency, naming and transparency, less storage space, high scalability, updated file replication and being the most important one is safe and secure access.Also the system requires minimal human administration and is less complex to use. Java Based Distributed File System provides the migration capabilities and location independency for data or information. The users are easily able to download and upload the files. Various problems that occur such as bad file replication, slow performance and poor scalability are resolved using Java Based Distributed File System. The system also enables effective mobile computing and file replication.

SYSTEM DEVELOPMENT METHODOLOGY AND TECHNOLOGY

Software Engineering is the process of practice of using selected process techniques to improvise on the quality of software development effort. This is on the basis of assumption that a methodical approach to software development result in fewer defects and therefore it ultimately provides us with shorter delivery times and better value. The challenge which one face in selecting and following a methodology is to do it judiciously. One system development methodology is not suitable for use. The methodology techniques that were selected for the development of system are Boehm Spiral Model, Rapid Application development and Prototype Model. Boehm Spiral model was selected for the advancement of the system. The spiral model uses incremental advancement, with the motive of managing risk. In the spiral model, advancers define and implement features in order of increasing priority. An initial version of the system is developed, and then it is repetitively modified on the basis of input received from evaluations. The advancement of each version of the system is carefully designed by the use of steps involved in. With each iteration around the spiral, progressively more complete version of the system is built. A key benefit of this approach is that the major issues are resolved at a very early stage. In the event that some of the difficulties cannot be resolved using one approach may be pursued. In the event that the difficulties cannot be resolved at all, the project requirements may need to be modified so that a time effective and feasible solution can be developed. This model was chosen because as mentioned above that spiral model has some benefits which are needed for the development of the system. The requirements remain that the end product should be capable enough to provide system with client server functionality in the form of packages of classes. Following the requirements gathering phase, an initial static analysis took place. The products of this analysis were the initial class diagrams of the core of the client package with the system for clients. During this phase, some of the class diagrams turned out to be defective, so they were plotted again. This procedure was followed several times for the client package in order to reach an acceptable state. The same technique was again used during the implementation of the remaining packages. In this way the idea of spiral model was used. The second methodology is prototyping. The methodology was selected to build the interface for client and interface [7]. A prototyping methodology is a software advancement process which permits the advancers to create portion of the solution to demonstrate functionality and make, needed refinements before advancing the end solution. The approach is favored as prototyping methodology is best suitable for implementing distributed systems, and can very quickly help in confirming system navigation and other user interaction requirements. It also allows the designer to receive feedback from the users at the earliest. The idea of Rapid Application Development was selected to implement the client-server model for Distributed File System. The obstacle facing software engineers and development organization can be summarized as more, better and faster. The RAD development paths access these obstacles on providing

system faster, while reducing cost and also by increasingly quality. The RAD methodology was chosen for this project because it also works when the work can be broken into manageable chunks like in this project. The advantage is also that the model work fine when the project is developed in small team.

The system development technology that has been adopted is Java as Java is a powerful platform for advancement of distributed systems. Java is one of the most popular programming languages used these days. Java provides us with all the necessary features for the advance network programming. The Distributed File System will be advanced on Java platform with Java's object-orientated and network-orientated application techniques. On the basis of conventional Distributed File System, the system will be applied on distinct operating environment, such as Windows, UNIX. Etc. The system also realises portability which is one of the most important characteristic of software engineering design. From installation to implementation, it ensures that users feel easy and comfortably during file operations. Extensibility and scalability also should be considered in the system. For this goal, the Java techniques will be illustrated in the design. Because of the apparent of simultaneously data passing, Java concurrency communication techniques application will be found in design process. Operational feasibility matches the requirements and expectations of users. To JDFS, user acceptance is a significant determination of operational feasibility. It requires careful consideration of. Corporate background knowledge meaning that JDFS can run on distinct operating system, so it is apt for distinct users who are familiar with distinct operating environment. Effective communication means user can access files stored in other computers as soon as his/her request is sent out. Vice versa, if other computers send request to the user, user's computer responds the request as soon as possible. Free tool kit is the system which is economical to be developed. Its design environment is on the basis of Java Runtime Environment which is free to download from Sun company website, so is J2SDK.Information is stored accurately in JDFS however the computer is on-line or not. The users can manipulate the files operation in a friendly user interface.Security is the most significant factors in system advancement. JDFS allocates username and passwords to each computer in the network. Accordingly, if users want to access the data shared in JDFS, they must log in certain specific computer by the authorised identity.

APPROACH TOWARDS DISTRIBUTED FILE SYSTEM

Software Interface Prototype for Client: In order to achieve the security, the file uploaded by the client on the server is encrypted by the server. The server then distributes those chunks of encrypted data among the clients. To get login with a login ID and a password, the client needs to get registered first. The admin here is the server itself. When the admin logs in, it takes the admin to the server window.
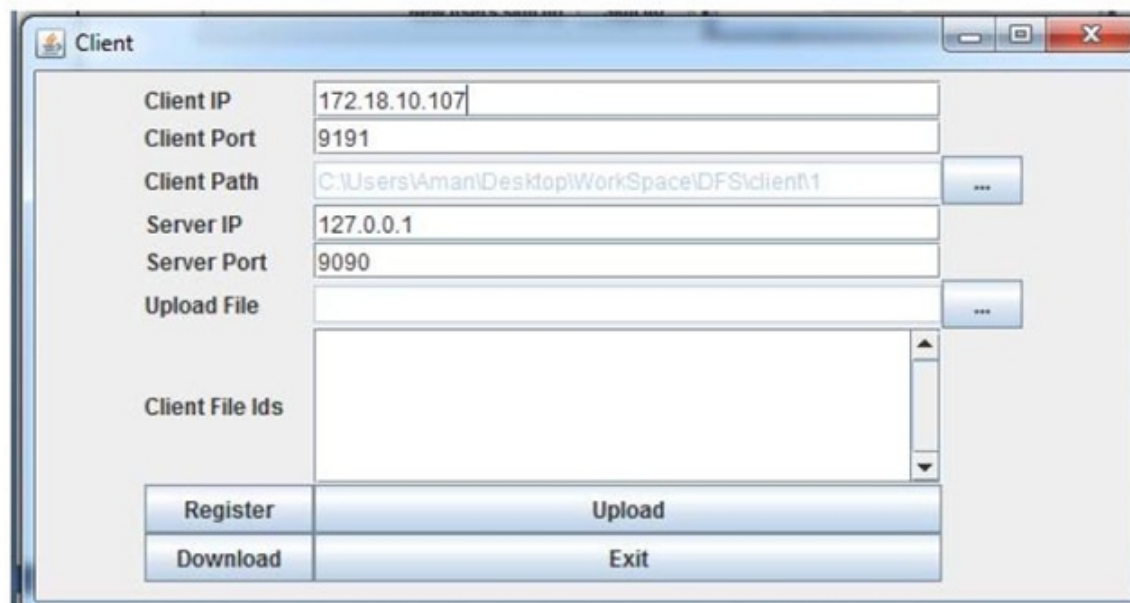
Fig1. Client Interface

This is the client window. The first text box is the Client's IP address which is fetched automatically and shown above. Then is the port number which should be different for all clients. Client Path is the path where the client wants to store the encrypted files in his/her system. Next is the Server IP address which the client needs to give once they log in which is this case is local host. Next is the Upload file text field where the client gives the path of the file which he wants to upload on the server. Client File Ids is the text area where the IP addresses and the port numbers of the entire client which get registered to the server is shown. This is just to let the client know that how many other clients are connected to the server at that time.

Software Interface Prototype for Server: As the server distributes the chunks of encrypted data among the clients which guarantees secure access, it also shares the key only with the client who requested to upload the file onto the server. Since the key is only shared between the client and server and the file is distributed in a random manner among the registered clients, there is no way for a hacker to get to know about the particular file. Now after the server sends the key to client, the original file can be downloaded. Once the key is selected by the user, the key merges those chunk sizes of data which were distributed among several clients back together and forms the original copy of that file[9]. Once the data is emerged and the file has been re-created, but in encrypted format, the key decrypts the file before sending it to the client who requested for that file download.
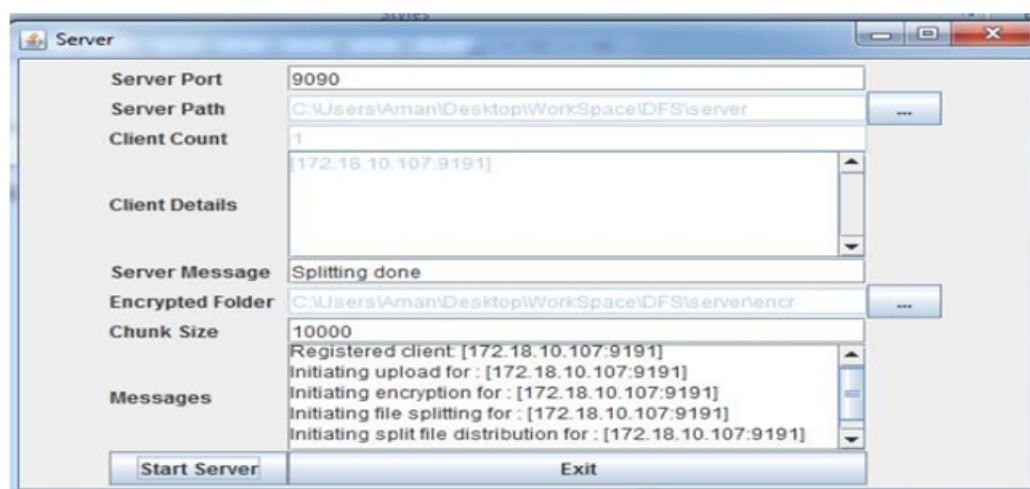
Fig. 2 Server Interface

When a file is uploaded by the client, the server encrypts the file, splits and distributes it to the different clients and sends a unique to the client who initiated the file upload. Once the key is selected the client clicks the Download button to download the file, the server merges all the chunk data from all the clients and merges them together and re forms the original file and decrypts it and sends it to the client. A message is shown in a message dialog box informing the client that the file has been decrypted and downloaded to that client.

CONCLUSION

The system has been successfully implemented. The system proved to be a new way in recovering all the problems that was occurring in the previous systems. The system has been successful in providing safe and secure access and enables effective communication. It also enables file recovery [11]. The system provides fault tolerance in case the CPU is more utilized; battery power has been more utilized etc.

The system provides file replication with updations that plays a vital role in case a client node does not work or becomes unavailable due to some situations. The naming and transparency of files has been successfully provided by the system. Various concepts have been solved and discussed using Distributed File System that has been implemented through Java. Semantics of file sharing and alternatives for the methods for providing access to remote files have been successfully provided by the system. Implementation of new approaches has been provided based on the assessment of other existing systems. The product has been critically examined and also possible improvement has been discussed.

REFERENCES

Park, G.L. (2004) . Performance Evaluation of a List Scheduling In Distributed Memory Multiprocessor System. International Journal of Future Generation Computer Systems, 20, 249-256.

Whitten, J., Bentley, L. & Dittman, K. (2004). Systems Analysis and Design Methods, 6th Ed, Mc-Graw-Hill.

R.J.T Morris Y.T.Wang. Load Sharing in Distributed Systems. IEEE Trans., Comput. C-34(3), March.1985.

Neuman, C.RFC 1510, "The Kerberos network authentication service (V5)", [S], 1993.

Roychoudhuri, L. & Al-Shaer E.S. (2005). Real-Time Packet Loss Prediction Based on End-to-end Delay Variation. IEEE Transactions On Network And Service Management, 2(1), 29-38.

Mirko K¨ampf, and Jan W. Kantelhardt. (2013). Hadoop.TS: Large-Scale Time-Series Processing. International Journal of Computer Applications, 74.

Gibson, R. & R. Van Meter, "Network Attached Storage Architecture. Communications of the ACM,13.

Tankel, D. (2010). Scalability of Hadoop Distributed File system, Yahoo developer work.

T.V Suresh Kumar, & S. Jagannatha, S. (2013) Algorithm of Performance Prediction by Resource Sharing in Distributed Database. International Journal of Computer Applications, 66.

Danilo Ardagnaa, Sara Casolari b, Michele Colajanni b, Barbara Panicucci a, Dual time -scale distributed capacity allocation and load redirect algorithms for cloud systems J. Parallel Distrib. Comput. 72 (2012) 796–808.

Dasgupta, Pallab, "Commit Protocols CS60002", Distributed Systems, Dept. of Computer Sc. &Engg, IIT Kharagpur.

Taubman, D.S. & M. W. Marcellin, M.W. (2002). JPEG2000", Image Compression, Fundamentals, Standards, and Practice by, Kluwer Academic Publishers.

PAIR PROGRAMMING: A Co-Creative Paradigm for Concerted Knowledge Acquisition and Competitive Advantage

[1]Dilip K Banerjee, [2]Supriya Gupta

[1]Central university of Jharkhand, Brambe, Ranchi 835205

[2]Central university of Jharkhand, Brambe, Ranchi 835205

2St. Xavier's College, Purulia Road, Ranchi 834001

1dkbanrg@gmail.comand2supriyagupta000@gmail.com

A B S T R A C T

Pair programming (PP) is one of the key practices of Extreme Programming which is a variant of Agile models. In PP two programmers try to solve a problem together sitting on the same workstation by sharing their ideas and communicating almost every moment. In this paper we try to show that PP isa particular case of Co-creation paradigm and is a good medium for Knowledge acquisition and transition within the IT organization or in training schools like educational institutes. This also creates an atmosphere for collaborative behavioral orientation andquicker innovation, resulting into Competitive Advantage for the organization.

Keywords- Agile Methods, Pair Programming, Co-creation, Knowledge Acquisition, Competitive Advantage

1. INTRODUCTION

Agile software development is a group of software development methods based on iterative and incremental development, where requirements and solutions evolve through collaboration between self-organizing, cross-functional teams. It promotes adaptive planning, evolutionary development and delivery, a time-bound iterative approach, and encourages rapid and flexible response to change. It is a conceptual framework that promotes foreseen interactions throughout the development cycle. The Agile Manifesto[1] introduced the term in 2001.

Pair programming is an agile software development technique in which two programmers work together at one workstation. One, the driver, writes code while the other, the observer, reviews each line of code as it is typed in[2]. The two programmers switch roles frequently. While reviewing, the observer also considers the strategic direction of the work, coming up with ideas for improvements and likely future problems to address. This frees the driver to focus all of his/her attention on the "tactical" aspects of completing the current task, using the observer as a safety net and guide. "None of us is as smart as all of us", which is the essence of "Co-creation", a term that was popularized by CK Prahalad and Venkat

Ramaswamy, (2000, 2004)[3,4] and Ramaswamy et al (2010)[5]. Co-creation involves working on new product and service ideas together with the customers who are going (at least you hope) to buy them.

Despite decades of research and development in Knowledge Management, there still exists a serious knowledge acquisition bottleneck. Neither traditional Artificial Intelligence based approaches, nor the more recent less formal Knowledge Management techniques been able to overcome the knowledge acquisition challenges[6]. In most of the universities and colleges teaching computerscience at graduate or at masters' level typically students are introducedto the technical and/or the managerial side of the computer science discipline whereas the social side remains vastly neglected. On the other hand when projects are shared among the students they not only share the responsibility but also their grades, the concept of collaboration and team building becomes an essential component of that learning process. The learning becomes improved as they learn through their mutual strengths and weaknesses. They learn quickly and do not commit the same errors committed by their partners. Similar practices also benefit the novices in software firms to acquire the knowledge of the peers or the more experienced ones working in the same organization.

The twenty-first century competitive landscape and the consequent organizational rivalry have increased the urge for a unique corporate paradigm which is causal in behavior and innovative in consequence. A causal approach calls for a higher degree of cohesiveness in the behavioral pattern of stakeholders. Most firms have the tendency to create knowledge and then disseminate it both down the line as well to the stakeholders through various channels and media. However, through this behavior a firm can only develop a capability or at times, unique capabilities but not necessarily a sustained ability to innovate. Such ability would come from the integration of knowledge into firm's strategy. Customers and other stakeholder in the contemporary business environment seem to be more empowered, well-informed and networked, and therefore, the autonomous creation of knowledge by the firm becomes a challenging issue. The concept of 'value co-creation' (Prahalad &Ramaswamy, 2004) goes a long way in justifying the strategic necessity of co-creation as a source of advantage or competitive advantage for the firms. It is to be duly acknowledged that the firm's strategic intents are no longer based on traditional sources of competitive advantage but on newer themes which characterize co-creation and innovation at all levels.

2. PAIR PROGRAMMING

Pair programming is a practice, whereby two programmers work side by side at the same computer, continuously collaborating on the same design, algorithm, code, or test[1,2], when the workers / students

working in pair feel responsible for their partner's success or failure . The two members of the team using the same computer for coding where one member acts as the driver who is actually the coder and the other member who is looking at the screen is the navigator or we can say the visionary focuses on fault finding and problem.Two students sharing the same computer communicate almost every minute which gives them better understanding of the problem and the courage to move in right direction.

Lucas Layman [7] has discussed pair programming with respect to students in one of his paper. According to him an increase in the organization of the students' and pairing them up will be beneficial when collaborating on complex tasks in other classes or in a professional career. Collaborative assignments made the students save time and provided them better insight of the project.

Nils Brede Moe, TorgeirDingsoyr and Tore Dyba[8]has discussed in his paper a team following a plan-driven model often consists of independently focused self-managing professionals , who resist change. A transition from high individual and low group autonomy to a high level of individual and group autonomy is probably the biggest challenge when introducing change-driven development based on self-organizing teams. Software development process depends significantly on team work, as does any process that involves human interaction. Two important factors to achieve team performance are feedback and communication.

Hamed Yaghoubi Shahir, Shervin Daneshpajouh, and Raman Ramsin[9] also have suggested that pair programming makes two people responsible for a task and they can cover each other's weaknesses. They suggest that light analysis and design approach which brings agile approach closer to traditional approach. Traditional approach harps on heavy design and documentation which is generally not adopted by the students, they want to quickly get over with design with very little documentation, so this light designmethod would be readily accepted by students.

Melnik and Maurer[10] in their paper report the results of a study on student perceptions using agile practices. The study involved forty-five student enrolled in three different academic programs (Diploma, Bachelor's and Master's) in two institutions to determine their perceptions of the use of extreme programming practices in doing their design and coding assignments. Overwhelmingly, students' experiences were positive to use the practices if allowed.

Laurie Williams , D. Scott McCrickard , Lucas Layman , Khaled Hussein[11]suggested   certain guidelines for students to work in pair. These guidelines are in the context of the HCI course undertaken by 22 students in the Fall 2007. The two guidelines are as follows. The first suggested that it is important

a common goal. The second additional guideline suggests that the mentor should refrain from telling the answers but instead point them in the right direction and encourage them to find answers together.

KirtiGarg and Vasudeva[12] also have suggested that it is very important to bridge the gap between industry and academic learning and thereby creating professionals who are well versed with theory and practice and have experienced the intricacies of real software development even before entering the industry. Software Engineering (SE) Educators have been advocating the use of non-conventional approaches for SE education for long. In this context, they conducted action-research to compare the effectiveness of a case study approach with conventional lecture based approach. Results indicated that case study approach is more beneficial which involves thought process of every member of the group. A level interactive communication starts which also helps students to strengthen the relationship among the peer members .Thus this learning-by-doing paradigm, specially the case study approach will be the right tool for creating an effective feedback and assessment method in such active learning.

Pearl Brereton, Mark Turner and RumjitKaur[13] in their paper mentioned studies on five aspects of a student w.r.t pair programming. These aspects are Exam marks/Assignment marks, Assignment Quality, Pass/success rate, Retention, Confidence and Enjoyment attitude .They found that there was effective improvement in the pass rates of undergraduate students, significant improvement in the retention rates of programming modules. There was a great sense of enjoyment and had a better command on the language. Improvement in quality of work and thereby increase in the level of confidence. All the studies showed positive reports except one study report that stated that pair programming had no significant effect on the students of non-Computer Science majors.

2.1 Potential Benefits of Pair Programming

Pair programming is an essential XP practice thatWilliams and Kessler [14] define as follows:"Pair programming is a style of programming inwhich two programmers work side by side at onecomputer, continually collaborating on the samedesign, algorithm, code, or test. One of the pair,calledthe driver, is typing at the computer or writing down adesign. The other partner, called the navigator, hasmany jobs, one of which is to observe the work of thedriver, looking for tactical and strategic defects.Tactical defects are syntax errors, typos, calling thewrong method, and so on. Strategic defects occur whenthe driver is headed down the wrong path—what isimplemented just won't accomplish what needs to beaccomplished. The navigator is the strategic, longrangethinker. Another great thing is that the driverand the navigator can brainstorm on-demand at anytime. An effective pair programming relationship isvery active. The driver and the navigatorcommunicateat least once every 45 to 60

seconds. It'salso very important to switch roles periodicallybetween the driver and the navigator."

When programming in pair, the drivers' work is constantly reviewed. This is called "Pair Review" [15]. This helps students to debug the problem faster and discussing their work with the peer member makes their tasks simpler. In this entire process the students achieve a higher level of satisfaction and the team members get to each other better which in turn helps them to do the task in a much better collaborated way. Students learn to be better team members i.e. they learn to be a part of team and work towards more professional culture.

Working in pair leads to exchange of knowledge and if they have any doubt or problem in their assignment they have the courage to ask to their respective mentor. This is called "Pair courage"[15].

Students who are less confident about any programming language tend to learn more from their partners, thereby having a stronger knowledge base. It's a constant learning process, which helps them evolve as a better working professional and having the right sense of soft skill which makes them more appropriate to work in an organization. Strong coders who are generally introverts and don't like working in a team, when paired with right person with proper guidance of the mentor tend to come out of their own shell.

2.2 Threats to Pair Programming

If the pairing is not right and there is difference of opinions then members tend to withdraw from the project and loose interest and focus, so it is very important to do the right pairing of students, thereby increasing the work of the mentor. Firstly the mentor has to understand the students' personality traits and academic soundness of each student then take decision of paring students. Secondly mentor has to constantly guide them, get effective feedback and then keep evaluating them. It is a time taking procedure not only for the mentor but also for the students. Mentor spends time understanding his/her students and constant evaluation. Students have to expend time on understanding his/her peer members. Students tend to rely on their peer members and tend to deviate from the path of self-learning. Students tend to regress to their old habits of slack and casual attitude if the constant pressure of the mentor is not there.

3. PAIR PROGRAMMING AND CO-CREATION

It is seen that the Pair Programming is in fact a particular case of Co-Creative paradigm where the collaboration is taking place between two programmers working simultaneously on the same

workstation, these two partners could be two students, two peers in the same firm or even one from IT organization who is the developer and the other is the user of the software application or product being developed. Such an approach is surely more sustainable than conventional systems of development as it has a self-correcting mechanism built-in the development method.

In applying co-creation, managers often wonder where to start. One can start by designing an engagement platform with some chosen individuals somewhere in the system (from inside and/or outside) and then evolve the design of the platform with the people who engage with it. The idea is to start simple and keep the design of the platform loose enough so that it can evolve as a function of people's observed engagement[5].

Devising an engagement platform (such as a website) involves experimentation and learning[16]; and usually several iterations and false starts. It is nearly impossible to build a co-creative platform that ensures participants a perfect experience right at the first time. Platforms evolve in trial-and-error increments, with the roles of co-creators often changing significantly over time. The fundamental point is: the design of engagement platforms evolves as a function of the co-creative process, through ongoing interactions among co-creators, expanding the space of experiences, the scope and scale of interactions, platform linkages and stakeholder relationships in the ecosystem. Companies on their way to co-creation may have to take additional steps to strengthen their engagement and collaboration with customers and other stakeholders, to keep interactions alive, and make the human experience the true focus of their co-creation. A natural tendency, even for organizations initially successful at co-creation, is to make their engagement platforms progressively more firm-centric and ignore the experience of customers and other participants that made them successful initially.
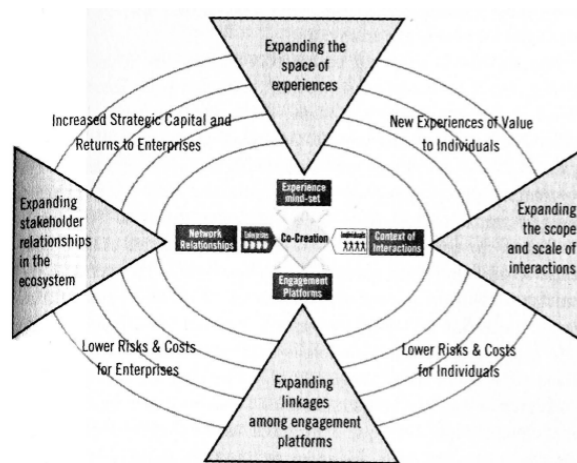

Figure 1: The Expanding Universe of Co-Creation Opportunities
(Taken from VenkatRamaswamy and Francis Gouillart (2010)- "The Power of Co-creation"[5])

4.KNOWLEDGE CO-CREATION AND INNOVATION ADVANTAGE

Firms that cannot continually engage in envisioning the need for innovation would have a difficult time in competing both in domestic market as well as internationally. Though the parameter for innovation may differ from market to market, yet the underlying philosophy remains the same.

Customers must perceive any innovation as being a high value offering made by the firm. Co-creation starts with the customers who view the market as a set of conversation between them and the firm.

Besides the behavioral aspect, technology also plays a pivotal role in sharing, co-creating and disseminating of knowledge across the firm and its stakeholders. Moreover, the roleof technology in the process of new product development also exemplifies its vitality in bestowing the firm its competitive edge to innovate. And in a market which is fast changing, innovation of course, is a strenuous task. It should not only focus on radical outcomes but also on continuous improvement of the existing offerings, so that the consequent newness brings in something which is distinct or unique.

Knowledge Co-creation (KC) may be defined as a strategic phenomenon which empowers the firm to continually enhance its learning capabilities by mutually sharing the knowledge which exists both at organizational levels as well as with stakeholders in order to create new knowledge thereby culminating the Innovation Advantage (IA).

The concept of knowledge co-creation, therefore, necessarily implies that:
• Knowledge Co-creation process is a strategic phenomenon as it empowers the firm.
• Learning capabilities can only be developed when both the firm and its stakeholders are able to share their existing knowledge and mutually disseminate it.
• Co-creation entails the emergence of new knowledge which is deemed to be an advantage-oriented strategy for the firm.
• By co-creating knowledge, the firm should focus on culminating the much desired Innovation Advantage (IA).

Innovation Advantage (IA) may be defined as the firm's sustained ability to outperform its competitors in terms of product, process, or structure, which should be its unique attribute and be newly induced into the system through the organizational efforts of Knowledge Co-creation (KC).

It implies that:

- Innovation Advantage is a strategic ability of the firm which strengthens its conviction to outperform its rivals in the highly competitive landscape.
- It is not a single step but a sustained ability to survive competitive rivalries.
- It should be attained either by the renewal of existing knowledge in a multidimensional perspective or by the creation of a new knowledge altogether.
- A segregated effort made by firms to create knowledge and then disseminate it across its stakeholders would not guarantee the desired Innovation Advantage.
- Innovation Advantage can necessarily be attained through cohesive efforts to co-create knowledge, as knowledge exists at all levels – both within and outside the firm.

4.1 The Innovation Advantage Matrix

The interplay of the two factors namely Behavioral Orientation and Technological Enticement is discussed through a framework known as 'Innovation Advantage Matrix[17] (IAM)'. The proposed matrix gives impetus to the co-creation approach for attainment of Innovation Advantage by the firms. The matrix comprises of four quadrants with two parameters viz. 'behavioral orientation' on the y-axis and 'technological enticement' on the x-axis.

4.2 Behavioral Orientation

The behavioral orientation amalgamates the firm's strategic priorities and helps it in identifying ways for generation of knowledge and enhancement of intellectual capital both within the firm and among its stakeholders. It needs to be emphasized herewith that the behavioral orientation is backed by an increased tendency among the firm and its stakeholders to share the knowledge. This sharing requires the appropriate technological enticement. The firm's behavioral orientation is composed of the following activities:

- Use of explicit knowledge for sustainability
- Management of implicit knowledge and development of firm's knowledge base. Prahalad and Ramaswamy[4] (2004) have also emphasized on co-creation of values through personalized interactions based on how each individual wants to interact with the company
- Intensity of stakeholders involvement and value creation
- Alignments of firm's learning capabilities with its strategic intents

4.3 Technological Enticement

The technological enticement of the firm helps the firm and its stakeholders in leveraging actionable knowledge by using the various information and communication tools. Using the networking tool for enhancing the sharing-
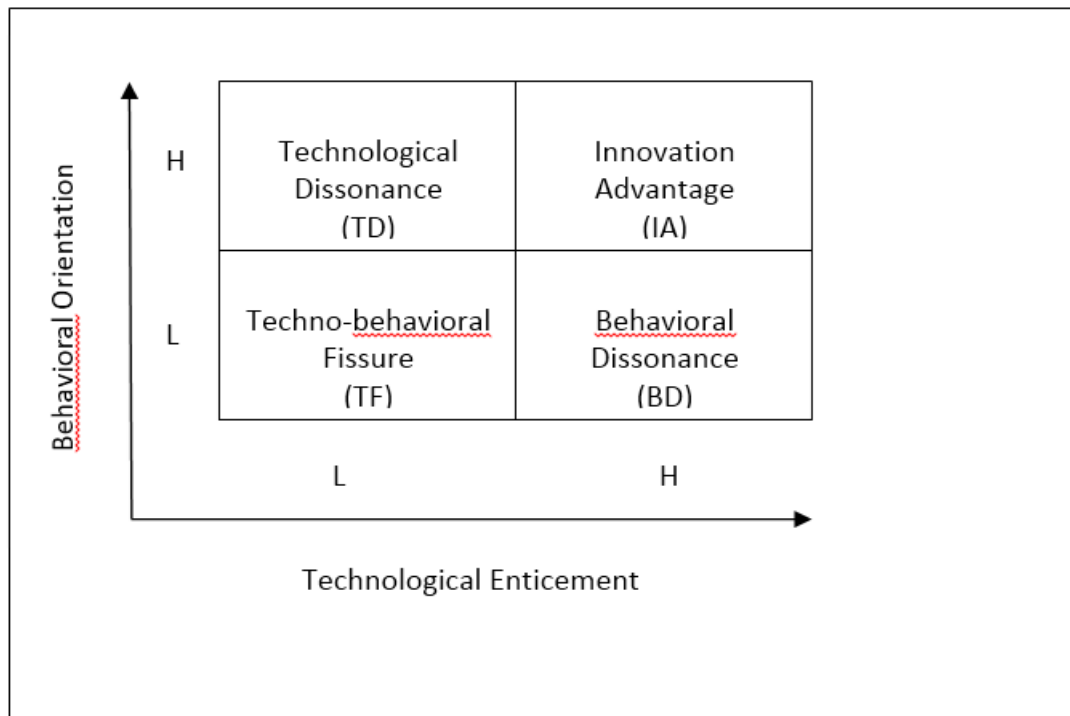

Figure 2: Innovation Advantage Matrix

effectiveness of co-created knowledge requires an efficient networking process. Efficiency in the process of sharing the existing knowledge requires an excellent networking across the firm and its stakeholders. The firm's technological enticement is composed of the following activities:

- Effectiveness in dissemination of co-created knowledge
- Propensity of new product design and development
- Product adaptation and standardization

The four different quadrants into which a firm could be classified based on the two parameters explained earlier are given below.

Techno-behavioral Fissure – A firm is said to be experiencing the state of techno-behavioral fissure if it witnesses a low in both the parameters of behavioral orientation and technological enticement.

It implies that lesser or negligible amount of knowledge is being created by the firm and its stakeholders; while, the technological advances for dissemination of knowledge among them is also very bleak or at times negligible. Co-creation does not happen at this stage. Firms which are either starters or small scale with a low turnover or those which includes distributors, etc. are normally classified in this quadrant. In order to move onto the next quadrant, such firms should follow any or all or any combination of the following strategies:

· Expand their area of operations

· Enhance their customer base

· Become a learning organization

· Enter into strategic alliances

Technological Dissonance – In this mode, the firms which are classified are high onbehavioral orientation but lack the appropriate and/or sufficient technological enticement for dissemination of the co-created knowledge. The knowledge co-creation process does take place in such firms but the lack of technical expertise and communication technologies creates a dissonance, which ultimately restricts its movement toward gaining the overall advantage in terms of innovativeness. In order to beclassified as one with Innovation Advantage, these firms should adopt any or all or any\combination of the following strategies:

· Develop technical expertise
· Use integrated communication technologies
· Reduce lead time
· Enter into strategic alliances for technology transfer

Behavioral Dissonance – The firm is said to be in a state of behavioral dissonance whenit is high on technological enticement but lacks the behavioral orientation. It implies thatthough such firms have sufficient technical expertise to make a move, yet the knowledgethat exists with them as well as with their stakeholder is not sufficient enough to gain thedesired innovation advantage. Such firms do co-create knowledge, though in a smallmagnitude. In order to be classified as one with the desired Innovation Advantage, thesefirms should adopt any or all or any combination of the following strategies:

· Enhance knowledge base within
· Encourage customers and other stakeholders to share their knowledge
· Seek more participation from customer and other stakeholders
· Enter into strategic alliances or outsourcing for enhancing their knowledge-based processes

Innovation Advantage – The firms falling under this quadrant are said to be bestowedwith the ability to cohesively integrate their knowledge and technical aspects within themas well as with their stakeholders to gain the Innovation Advantage. This is the highestlevel in the matrix, which strategically justifies the positioning of a firm and its ability toinnovate with respect to its competitors – both within the industry and outside. Suchfirms are capable enough to co-create knowledge and sustain their ability to innovatewith the changing competitive landscape. Such firms should use a strategy that enablesthem to retain their competitive advantage and sustain their ability to co-create andinnovate. Such firms should also enhance their goodwill by defining their newer targetof corporate social responsibility.

5. CONCLUSION

Pair Programming is a great learning process which helps knowledge acquisition and transmission among the stake holders. If the client/user (or its representative) performs the reviewer's role in the development process on a continuous basis along with the developer or coder then the chance of failure of the project is minimized. At the same time it offers scope for higher degree understanding on both sides, the coder learns the nitty gritty of the users' business process and the client too understands that it may be difficult or impossible to implement everything that is initially perceived of the system. This co-created new knowledge may be leveraged by the software firm to educate the human capital they employ and also helps better realize the potential of Information Technology as enabler for the users.

Often through knowledge co-creation new Innovationsare generated. Such Innovation Advantage could be made to be arepetitive phenomena and not merely one time lucky event. It is imperative that firms mustcomprehend that knowledge multiplies with experience and with each repetition itstransformation to next higher actionable level takes place. The paper aims to incorporateall 'stakeholders' instead of 'customers' alone in the process of knowledge co-creationfor attaining the desired Innovation Advantage.

Finally as teaching and training tool Pair Programming may open up new vistas for active learning with peers which is faster, deeper and unleashing unbound energy and enjoyment of achievement through team work. However, the role of the mentor in this process is vital and cannot be over emphasized.

6. REFERENCES

[1] Beck, Kent; et al. (2001). "Manifesto for Agile Software Development", Agile Alliance, http://agilemanifesto.org/.

[2] *Williams, Laurie (2001).* *"Integrating Pair Programming into a Software Development Process"* *(PDF). 14th Conference on Software Engineering Education and Training: abstract.* *http://ieeexplore.ieee.org/Xplore/login.jsp?url=/iel5/7294/19720/00913816.pdf.*

[3] Prahalad, C.K.; Ramaswamy, V. (Jan.-Feb. 2000) "Co-Opting Customer Experience".Harvard Business Review.

[4] Prahalad, C.K.; Ramaswamy, Venkat (2004) The Future of Competition. Harvard Business School Press.ISBN 1-57851-953-5.

[5] Ramaswamy, Venkat; Gouillart, Francis (2010). The Power of Co-Creation: Build It with Them To Boost Growth, Productivity, and Profits. Free Press.

[6] Christian Wagner (2004), "Collaborative Knowledge Management: Breaking the Knowledge Acquisition Bottleneck", published in the 'Innovation Through Information Technology' (IRMA International Conference 2004).

[7] Lucas Layman ,''Changing Students' Perception: An Analysis of the supplementary Benefits of Collaborative Software Development, North Carolina State University,2006.

[8] Nils Brede Moe, TorgeirDingsoyr , Tore dyba, " Understanding Self Organizing Teams in Agile Software Development'',SINTEEICT,Norway, 2008.

[9]HamedYaghoubiShahir, ShervinDaneshpajouh, and Raman Ramsin, "Improvement Strategies for Agile Processes: A SWOT Analysis Approach", Department of Computer Engineering Sharif University of Technology Tehran, Iran, 2008.

[10] Melnik, G., Maurer, F., 2002. Perceptions of Agile Practices: A Student Survey, A student survey —Extreme Programming and Agile Methods_XP/Agile Universe 2002, Vol. 2418, pp. 103-113.

[11] Laurie Williams , D. Scott McCrickard , Lucas Layman , Khaled Hussein, "Eleven Guidelines for Implementing pair Programming in the Classroom", North Carolina State University ,Department of Computer Science,2008.

[12] Garg, K., Varma, V., 2007. A Study of the Effectiveness of Case Study approach in Software Engineering Education, in: Proceedings of 20th. Conference on Software Engineering Education & Training (CSEET'07), July 3-5, pp. 309-316.

[13] Pearl Brereton, Mark Turner,RumjitKaur,''Pair Programming as a teaching tool: a student review of empirical studies'' ,School of Computing and Mathematics,2009

[14] L.Williams and R.Kessler, Pair Programming Illuminated,Addison Wesley, Reading, Massachusetts, 2003

[15] Chih-weiHo, Kelli Slaten, Laurie Williams and Sarah Berenson, "Unexpected Student Outcomefrom Collaborative Agile Software DevelopmentPractices and Paired Programming in a SoftwareEngineering Course",2004.

[16] DilipKumarBanerjee, Co-creating in the Cloud, National Conference on Contemporary Issues in Business, Management & Finance (CIBMF-2012) held in University of Kashmir, 21st Sept – 23rd Sept, 2012.

[17] Faisal Ahmed and Dilip K. Banerjee (2008), Co-creating Knowledge for Innovation Advantage: A Strategic Necessity,Annual International Conference of Strategic Management Society, USA (SMS 2008) held in Indian School of Business (ISB), Hyderabad, December 12-14, 2008.

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

-

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).

2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);

3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);

4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);

2. Informative contribution (editorial, commentary, etc.);

3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.

The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

Address of the Editorial Office:

Enriched Publications Pvt. Ltd.
S-9,IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR -5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-45525005

Notes: